A Bayesian hierarchical spatio-temporal model for extreme temperatures in Extremadura (Spain) simulated by a Regional Climate Model

José Agustín García[1,2] · Francisco Javier Acero[1,2] · Javier Portero[1]

Abstract

A statistical study was made of the temporal trend in extreme temperatures in the region of Extremadura (Spain) during the period 1981–2015 using a Regional Climate Model. For this purpose, a Weather Research and Forecasting (WRF) Regional Climate Model extreme temperature dataset was obtained. This dataset was then subjected to a statistical study using a Bayesian hierarchical spatio-temporal model with a Generalized Extreme Value (GEV) parametrization of the extreme data. The Bayesian model was implemented in a Markov chain Monte Carlo framework that allows the posterior distribution of the parameters that intervene in the model to be estimated. The role of the altitude dependence of the temperature was considered in the proposed model. The results for the spatial-trend parameter lend confidence to the model since they are consistent with the dry adiabatic gradient. Furthermore, the statistical model showed a slight negative trend for the location parameter. This unexpected result may be due to the internal and modeling uncertainties in the WRF model. The shape parameter was negative, meaning that there is an upper bound for extreme temperatures in the model.

Keywords Extreme temperature · Regional climate model · GEV distribution · Bayesian hierarchical model

1 Introduction

It has now become quite clear that the emission into the atmosphere of huge amounts of CO_2 from the combustion of fossil fuels by humankind, together with other changes such as deforestation and intensive agriculture, is altering the climate, with rising temperatures over most of the planet (Cubasch et al. 2013). One concern for society and

its policymakers is that higher average temperatures may enhance the risk of extreme events such as heatwaves. These are associated with effects that can be lethal for humans, and, even below lethality, can have adverse impacts on their livelihoods (Field et al. 2012). Examples of such heatwaves were those that occurred in Southern Europe in 2003 (García-Herrera et al. 2010), Russia in 2010 (Barriopedro et al. 2011), and Canada in 2021 Thompson et al. (2022).

The best tools with which to project the future evolution of these extreme events are global climate models (GCMs) that take account of forcing from greenhouse gas emissions (see for example (Randall et al. 2007)). At present however, these models are too coarse to properly represent atmospheric properties such as temperature or precipitation at the local scale as would be needed for adaptive planning and policy making. This issue is usually addressed by downscaling the GCMs using Regional Climate Models (RCMs) (see for example (Xu et al. 2019; Tapiador et al. 2020) and references therein). An RCM is a limited-area weather model embedded in a GCM, or a higher resolution reanalysis that takes inhomogeneities of the Earth's surface into account as well as other properties not covered by the coarser GCMs. Obviously, if an RCM is to be used to study the future state

José Agustín García, Francisco Javier Acero and Javier Portero have contributed equally to this work.

✉ Francisco Javier Acero
fjacero@unex.es

José Agustín García
agustin@unex.es

Javier Portero
jportero@unex.es

[1] Departamento de Física, Universidad de Extremadura, Avd. Elvas s/n, Badajoz 06006, Spain

[2] Instituto Universitario de Investigación del Agua, Cambio Climático y Sostenibilidad (IACYS), Universidad de Extremadura, Avd. Elvas s/n, Badajoz 06006, Spain

of the climate then it must describe reasonably well the current state. The variable on which the present study will focus is that of extreme temperatures.

There are two main ways to approach the study of extreme meteorological and climate events. One is to use extreme indices such as some high percentile of the probability distribution function (Zhang et al. 2010; Peterson et al. 2001; Peterson 2005). The other is to apply the branch of statistics known as extreme value theory (EVT). An important advantage of the latter is that the technique is more informative than that of the extreme indices since, by obtaining the parameters of the distribution function, one can estimate the likelihood of as-yet unobserved events—the 100-year return level for example. This feature is widely used for engineering design purposes. The main drawback of EVT is probably the requirement for independence of the data which is needed to estimate the parameters of the distribution. But Coles (2001), among others, have shown that under conditions of weak dependence between the variables, the EVT can still be applied. The present work in particular applies the GEV formulation of EVT in order to get a climatological picture of extreme temperature events through estimating the parameters of the corresponding extreme value distribution. Given the spatial character of temperature as a variable, it seemed appropriate to use the theory of spatial extremes to study it. Three approaches have typically been used to address the problem of spatial extremes—max-stable random fields, copulas, and Bayesian hierarchical models (Davison et al. 2012)—and it is this last approach we shall take in the present work. One of the more important benefits of using a spatial extremes theory instead of modeling each observatory individually is the increased precision when estimating the parameters of the statistical distribution. Indeed, one of the difficulties one faces in a statistical study of extreme data is that, by their very nature, extreme data are 'rare', i.e., there are usually so few data that one must expect to get a large uncertainty in the parameters that describe the tail of the statistical distribution. One way to mitigate the consequences of this lack of data and to increase the precision of the parameter estimates is to trade space for time, pooling information from different observatories (see, for example, (Casson and Coles 1999; Cooley et al. 2007; Schliep et al. 2010)). Also, as pointed out by Renard (Renard 2011), a spatial theory allows the parameters of the extreme distribution to be estimated at an ungauged or poorly gauged site. This important advantage allows one to address the so-called change of support problem in spatial statistics (see (Gelfand et al. 2001; Fuentes et al. 2003)). This problem arises when one tries to compare data coming from sources at different spatial scales. In our case, we want to compare the extreme value distribution obtained from a model defined at a grid-cell scale with that obtained at a meteorological observatory defined at a local scale. Moreover, the use of Bayesian

statistics allows one to account properly for the uncertainties that naturally arise when modeling meteorological phenomena (see Epstein (1985)).

Most published studies dealing with extreme weather and climate events obtained from GCM/RCM models used extreme indices (Sillmann et al. 2013a, b; Bartolomeu et al. 2016; Deng et al. 2016; Jiang et al. 2015; Lorenz et al. 2016; Zollo et al. 2016). Far fewer have used EVT. As one example of applying EVT, Kharin and Zwiers (2005) made a study of extreme temperature and precipitation in a GCM for the period 1990–2100, considering several scenarios using a GEV distribution. Also, Tomassini and Jacob (2009) made an analysis of extreme precipitation in Germany using a non-homogeneous Poisson point process to determine the GEV parameters. Both cases, however, fitted the statistical model to the results obtained at each grid point individually. On the other hand, Schliep et al. (2010) used a spatial hierarchical model in a study of extreme precipitation obtained from six different RCMs. That study used a multivariate intrinsic autoregressive (IAR) spatial model—a kind of Gaussian Markov Random Field model. While this type of model is computationally well suited to fitting the gridded output of a climate model, it is less suitable for making a prediction at a 'non-observed' site. A model closer to the one used in the present work is that of Craigmile and Guttorp (2013) who studied extreme minimum temperatures obtained in Sweden from an RCM. The main problem we have found in that paper is that the analysis does not take the altitude dependence of the temperature into account, a dependence that is now included in the present model.

The objective of the present study was therefore two-fold—firstly, to simulate the present climate in our region by means of an RCM, specifically, the WRF v4.0 model using ERA Interim reanalysis boundary conditions, and secondly, to produce a statistical model which can be used to characterize summer extreme temperatures produced by the RCM for the region of Extremadura (Spain). This statistical model will then be used to assess whether the RCM can correctly describe the region's observed extreme temperatures.

The paper is organized as follows. The data used are described in Sect. 2, the details of the statistical model are presented in Sect. 3, and the results in Sect. 4. Finally, some conclusions are drawn and discussed in Sect. 5.

2 Data

The temperature data used come from a simulation of the climate using the community WRF model (version 4.0.1) (Skamarock et al. 2008) forced with ERA Interim (Dee et al. 2011) reanalysis. The period for the simulation was 1980–2015. The simulation used two two-way nested domains centred in the Iberian Peninsula. The larger external
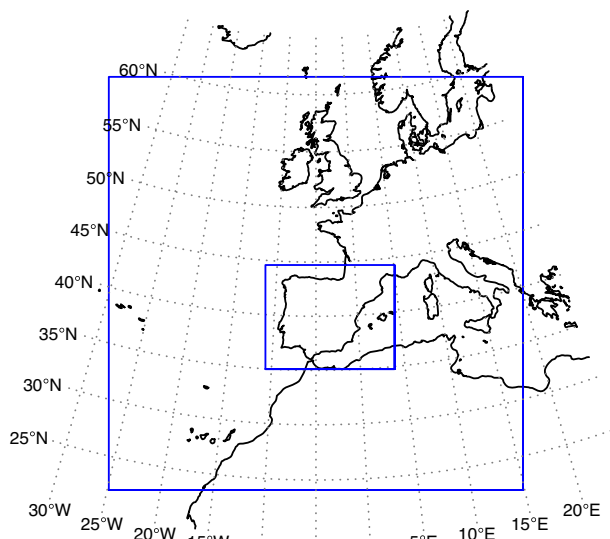
Fig. 1 The blue boxes in the map shows the position of the two domains used in the simulation. The larger external domain's resolution was 36 km, and the internal one's was 9 km

domain's resolution was 36 km, and the internal one's was 9 km. Figure 1 shows the position of the two domains. The number of vertical levels is 50.

Every 6 h, the boundary conditions, including Sea Surface Temperature (SST) and deep-soil temperature updates, were provided, and analysis nudging was applied in the external domain beyond the Planetary Boundary Layer (PBL). The WRF physical configuration was: Microphysics—WRF Single-Moment 6-class (Hong and Lim 2006); Longwave/Shortwave Radiation—RRTMG (Iacono et al. 2008); Surface Layer—MM5 similarity (Jiménez et al. 2012); Planetary Boundary Layer—Yonsei University (Hong et al. 2006; Hong 2010); Land Surface—Noah Land Surface Model (Chen and Dudhia 2001); Cumulus

Parametrization—Grell-Freitas (Grell and Freitas 2014). The model was run at intervals of 6 years, disregarding the first as a spin-up, and combining the results at the end. The version of WRF used was that modified by (Fita et al. 2019) which allows the user to obtain extreme values from the data simulated at each step of the modeling process.

In the inner domain, we focused on the region of Extremadura (Fig. 2).

For comparison with the results obtained with our statistical model, a set of extreme temperatures in the period 1981–2015 observed at 28 meteorological observatories sited in the aforecited region of Extremadura was used. Table 1 lists the code, name, and geographic coordinates of these observatories. Figure 3 shows their positions within the region of Extremadura.

For both the WRF model and the observed data, the extreme temperature is the highest temperature during the summer season (June–July–August).

We also used the Stead database developed by Serrano-Notivoli, Begueria, and De Luis Serrano-Notivoli et al. (2019) to compare the climatology of the maximum daily summer (June–July–August) temperatures given by the WRF model.

3 Statistical model

As was noted in the Introduction, we shall use a Bayesian hierarchical model framework in which to address the problem of estimating the parameters of a spatial Extreme Value model. In this framework (see Berliner (2003); Cressie (2011)), in the first stage, denoted 'data model', it is assumed that the observed noisy data are random variables that depend on a latent (unknown) process plus various parameters. In the second stage, denoted 'process model', a model is proposed for the latent process through a conditional

Fig. 2 Location of the Extremadura region in the Iberian Peninsula (left). Topographic map of Extremadura as seen by the WRF model (right) with the scale in the colorbar in metres above sea level (m.a.s.l.)
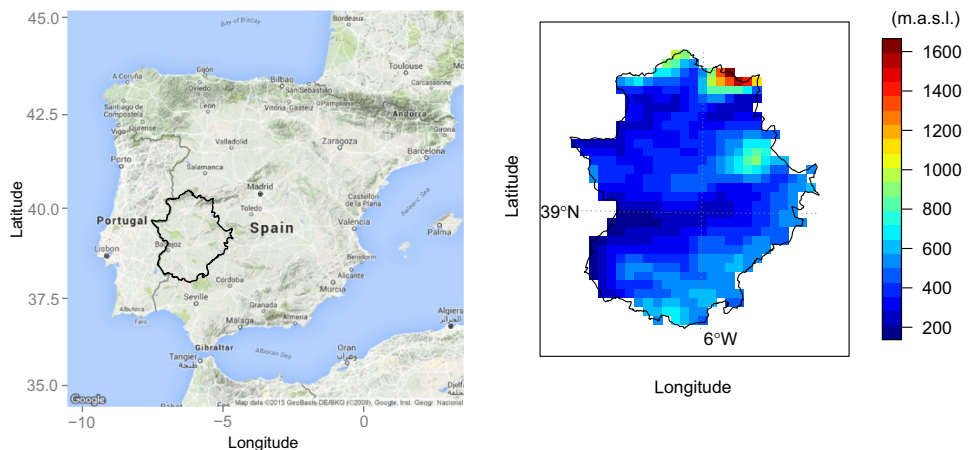
Table 1 Code, name, and geographical coordinates of the meteorological observatories used in the study

Observatory code	Name	Longitude	Latitude
3384	GARVIN DE LA JARA	−5.358	39.689
3429	VILLANUEVA DE LA VERA	−5.382	40.078
3439	BARRADO	−5.882	40.083
3448	SERRADILLA	−6.141	39.830
3455	ALDEACENTENERA	−5.678	39.568
3469	CACERES	−6.338	39.471
3502	PANTANO GABRIEL Y GALAN	−6.125	40.220
3528	CORIA	−6.590	39.980
3531	VILLANUEVA DE LA SIERRA	−6.404	40.200
3562	HERRERUELA	−6.906	39.462
3575	SANVICENTE ALCANTARA	−7.136	39.362
3576	VALENCIA DE ALCANTARA	−7.247	39.416
4244	HERRERA DEL DUQUE	−5.049	39.165
4245	GUADALUPE	−5.333	39.455
4320	PANTANO DE ZUJAR	−5.485	38.916
4358	DON BENITO	−5.860	38.955
4385	VALENCIA DE LAS TORRES	−5.985	38.447
4411	ALCUESCAR	−6.228	39.180
4429	LOS SANTOS DE MAIMONA	−6.382	38.450
4436	ALMENDRALEJO	−6.361	38.690
4444	MONTIJO	−6.615	38.911
4452	BADAJOZ/TALAVERA LA REAL	−6.813	38.883
4471	PRESA PEÑA DEL AGUILA	−6.876	39.140
4484	PANTANO DE PIEDRA AGUDA	−7.017	38.687
4500	CALZADILLA DE LOS BARROS	−6.316	38.300
4511	JEREZ DE LOS CABALLEROS	−6.771	38.318
4520	FREGENAL DE LA SIERRA	−6.651	38.168
5769	MONTEMOLIN	−6.136	37.994

probability distribution. It is in this second stage that scientific models are usually introduced. The third stage, denoted 'parameter model', can be considered the Bayesian part of the hierarchical model. In it, the parameters are considered random variables with a probability distribution function to take into account the uncertainties about them.

Stage 1 (data model) $\mathbf{P}(\text{data} \mid \text{process}, \text{parameters})$

Stage 2 (process model) $\mathbf{P}(\text{process} \mid \text{parameters})$

Stage 3 (parameter model) $\mathbf{P}(\text{parameters})$

Bayes' theorem allows one to determine the distribution of the model's parameters once the data have been observed. The corresponding relationship is
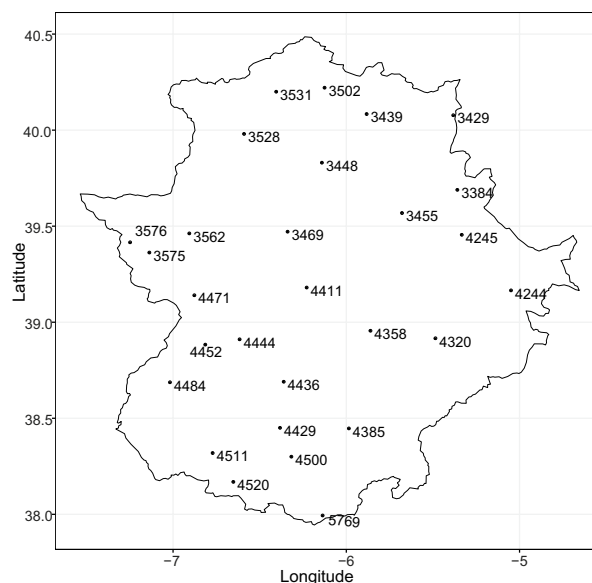

Fig. 3 Location within the Extremadura region of the meteorological observatories

$$\mathbf{P}(\text{process, parameters} \mid \text{data})$$
$$\propto \mathbf{P}(\text{data} \mid \text{process, parameters}) \cdot \mathbf{P}(\text{process} \mid \text{parameters})$$
$$\cdot \mathbf{P}(\text{parameters}).$$
$$(1)$$

The distribution on the left hand side of Equation (1) is termed the *posterior* distribution, the distribution $\mathbf{P}(\text{parameters})$ on the far right the *prior* distribution, and the distribution $\mathbf{P}(\text{data}|\text{process}, \text{parameters})$ in the middle the *likelihood*. One of the problems with Equation (1) is that the proportionality constant is unknown. Its evaluation is only feasible in simple cases. The introduction of numerical techniques such as Markov chain Monte Carlo (MCMC) methods (Gilks et al. (1996)), however, has allowed numerical simulation of the posterior distribution of the parameters, in particular by obtaining a sample of the parameters of interest. In the following subsections, we shall describe the proposed model in greater detail.

This hierarchical model may be placed in a more specifically climatological context (see for example Berliner (2003), Cooley et al. (2007)). It is well known that the weather, i.e., the state of the atmosphere at a given time and place, may be regarded as a random process and the relevant weather variables (pressure, temperature, etc) as random variables. The values of these random variables are expressed by probability distributions whose parameters may be considered to represent the climate of that place. This process corresponds to Stage 1 of the hierarchical model. Furthermore, in Stage 2, the climate, represented by the parameters of the distribution, is

itself regarded as a random process expressed by a probability distribution. Finally, in Stage 3, a probability distribution is proposed for the prior of the parameters that appear in Stage 2.

3.1 First stage (data model)

In the first stage, we assume that the block extreme data at observatory s, Y_s follow a GEV distribution, where the probability distribution function can be expressed as

$$\mathbf{P}(Y_s \leq y | \mu_s, \sigma_s, \xi_s) = \exp\left\{-\left[1 + \xi_s\left(\frac{y - \mu_s}{\sigma_s}\right)\right]^{-1/\xi_s}\right\},$$
$$1 + \xi_s\left(\frac{y - \mu_s}{\sigma_s}\right) > 0. \tag{2}$$

In the present study, the block extreme data are maximum temperatures in the summer season (June-July-August). It is also assumed that, conditional on μ_s, σ_s and ξ_s, the block extremes are spatially and temporally conditionally independent, i.e., the block extreme at observatory s conditional on μ_s, σ_s, and ξ_s is independent of the block extreme at observatory s', and that the block extreme observed at time t is independent of that observed at time t'. The scale parameter σ_s represents the spread of the distribution, with a greater value representing greater dispersion of the maximum temperature. The shape parameter ξ_s represents the tail behaviour of the extreme distribution—a negative parameter corresponds to a bounded tail, otherwise the tail is unbounded.

3.2 Second stage (process model)

In the second stage, the parameters of the GEV distribution are assumed to follow a spatio-temporal model, i.e., to depend on spatial coordinates $\mathbf{s}$ and time t. At this point, it is important to distinguish between stationary and non-stationary models. In the former, the GEV parameters do not depend on time, while in the latter they do. In the stationary case, the parameters of the GEV distributions are assumed to take the form

$$\mu(\mathbf{s}) = X_\mu(\mathbf{s}) \cdot \boldsymbol{\alpha}_\mu + W_\mu(\mathbf{s}) + \epsilon_\mu \tag{3}$$

$$\sigma(\mathbf{s}) = X_\sigma(\mathbf{s}) \cdot \boldsymbol{\alpha}_\sigma + W_\sigma(\mathbf{s}) + \epsilon_\sigma \tag{4}$$

$$\xi(\mathbf{s}) = X_\xi(\mathbf{s}) \cdot \boldsymbol{\alpha}_\xi + W_\xi(\mathbf{s}) + \epsilon_\xi \tag{5}$$

where $X_{\mu,\sigma,\xi}(\mathbf{s})$ are p spatial covariates (geographical coordinates) that may be different for each parameter, $\boldsymbol{\alpha}_{\mu,\sigma,\xi}$ is a set of p regression parameters, $W_{\mu,\sigma,\xi}(\mathbf{s})$ are spatial models

capturing the associations between different sites (grid cells), and $\epsilon_{\mu,\sigma,\xi}$ represent noise unaccounted for in the spatial models (generally known in the spatial data community as the nugget effect), and $\mathbf{s}$ denotes a spatial coordinate. In the non-stationary case, a linear time dependence is included in the location parameter:

$$\mu(\mathbf{s}, t) = X_\mu(\mathbf{s}) \cdot \boldsymbol{\alpha}_\mu + W_\mu(\mathbf{s}) + (X'_\mu(\mathbf{s}) \cdot \boldsymbol{\gamma}_\mu$$
$$+ W'_\mu + \epsilon'_\mu)(t - t_0) + \epsilon_\mu. \tag{6}$$

This equation can be rewritten as

$$\mu(\mathbf{s}, t) = \mu_0(\mathbf{s}) + \mu_1(\mathbf{s})(t - t_0). \tag{7}$$

In Eq. (6), it has been assumed that the temporal-trend coefficient may depend on geographical factors through the covariates X'_μ, and that nearer sites could show a stronger temporal relationship through the spatial model W'_μ. The spatial models $W_\kappa(\mathbf{s}), \kappa = \mu, \sigma, \xi; W'_\mu(\mathbf{s})$ are assumed to be of the form

$$W_\kappa(\mathbf{s}) = Z\Psi_\kappa(\mathbf{s}'); \qquad \kappa = \mu, \sigma, \xi \tag{8}$$

where Z is an $n \times r$ matrix and $\Psi_\kappa(\mathbf{s}')$ is a spatial process of dimension r. This spatial process is modeled as a multivariate Gaussian process with zero mean and covariance matrix $\boldsymbol{\Sigma}$, i.e.,

$$\Psi_\kappa(\mathbf{s}') \sim MVN(\mathbf{0}; \boldsymbol{\Sigma}_\kappa); \qquad \kappa = \mu, \sigma, \xi \tag{9}$$

The $r \times r$ covariance matrix defines the covariance among the r spatial locations, and is assumed to be of the form

$$\boldsymbol{\Sigma}_\kappa(s', s') = \beta_{\kappa 1}^2 \exp(-\|s_i' - s_j'\|)/\beta_{\kappa 2} = \beta_{\kappa 1}^2 R_\kappa(\beta_{\kappa 2}) \tag{10}$$

where R_κ is the correlation matrix. The parameter $\beta_{\kappa 2}$ (also called the range parameter) defines the strength of the association in the relationships among the spatial sites. The variance parameter $\beta_{\kappa 1}^2$ is also known as the sill in the geostatistics community. The residual noise ϵ_κ is modeled as a Gaussian process of zero mean and variance τ_κ^2 which is assumed to be constant everywhere.

The reason for the introduction into our model of a matrix Z that transforms an r-spatial process into an n-spatial process is the following. The number of grid cells increases enormously as the resolution of the model increases. For example, for our region of Extremadura and with a WRF resolution of 9 km, there are 957 grid cells. Such a large number of cells makes the numerical problem nearly unfeasible (the covariance matrix would be 957×957 and we would have to invert such a large matrix at each MCMC step). To lighten the numerical burden, we followed the method proposed in Banerjee et al. (2008), and modified in Finley et al. (2009) (see also Eidsvik et al. (2012)). Those authors

proposed replacing the W_κ process by what they called a *predictive process* $\tilde{W}_\kappa$, defined as

$$\tilde{W}_\kappa(s) = E(W_\kappa(s)|\Psi_\kappa(s')) = \Sigma_\kappa(s,s') \cdot \Sigma_\kappa^{-1}(s',s')$$
$$\Psi_\kappa(s') = Z_\kappa \Psi_\kappa(s'), \tag{11}$$

where $\Sigma_\kappa(s,s')$ is the $n \times r$ covariance matrix between an **s**-site and an **s**'-site. With a value $r \ll n$, the problem of inverting an $n \times n$ matrix at each MCMC step reduces to that of inverting an $r \times r$ matrix. In this work, we have sub-sampled the WRF grid every three rows and columns, which yields an r-matrix of 110×110 elements. The substitution of the W_κ process by the predictive $\tilde{W}_\kappa$ process has the consequence of reducing the variance of the spatial process. To compensate for this, Finley et al. (2009) proposes increasing the nugget variance $\tau^2 I$ by the amount

$$\text{Diagonal}(I - \Sigma_\kappa(s,s')\Sigma_\kappa^{-1}(s',s')\Sigma_\kappa(s',s)).$$

To close this second stage, Equation (3) may be expressed as normal distributions: for the stationary model,

$$\mu(\mathbf{s}) \sim N(\mathbf{X}_\mu(\mathbf{s}) \cdot \boldsymbol{\alpha}_\mu + W_\mu(\mathbf{s}), \hat{\tau}_\mu^2)$$
$$\sigma(\mathbf{s}) \sim N(\mathbf{X}_\sigma(\mathbf{s}) \cdot \boldsymbol{\alpha}_\sigma + W_\sigma(\mathbf{s}), \hat{\tau}_\sigma^2)$$
$$\xi(\mathbf{s}) \sim N(\mathbf{X}_\xi(\mathbf{s}) \cdot \boldsymbol{\alpha}_\xi + W_\xi(\mathbf{s}), \hat{\tau}_\xi^2),$$

and for the non-stationary model,

$$\mu_1(\mathbf{s}) \sim N(\mathbf{X}_\mu(\mathbf{s}) \cdot \boldsymbol{\alpha}_\mu + W_\mu(\mathbf{s}), \hat{\tau}_\mu^2)$$
$$\mu_2(\mathbf{s}) \sim N(\mathbf{X}'_\mu(\mathbf{s}) \cdot \boldsymbol{\gamma}_\mu + W'_\mu(\mathbf{s}), \hat{\tau}'^2_\mu)$$
$$\sigma(\mathbf{s}) \sim N(\mathbf{X}_\sigma(\mathbf{s}) \cdot \boldsymbol{\alpha}_\sigma + W_\sigma(\mathbf{s}), \hat{\tau}_\sigma^2)$$
$$\xi(\mathbf{s}) \sim N(\mathbf{X}_\xi(\mathbf{s}) \cdot \boldsymbol{\alpha}_\xi + W_\xi(\mathbf{s}), \hat{\tau}_\xi^2),$$

where the $\hat{\tau}_i$ represent the increased nugget variances.

3.3 Third stage (prior model)

In the third stage, prior distributions have to be provided for the parameters used in the previous two stages, in particular, for the spatial regression coefficients $\boldsymbol{\alpha}_\kappa$, the parameters of the covariance model $\beta_{\kappa 1}^2, \beta_{\kappa 2}$, the nugget variances τ_κ^2, and, in the non-stationary case, the trend parameters $\boldsymbol{\gamma}_\kappa, \beta_{\kappa 1}'^2, \beta_{\kappa 2}', \tau_\kappa'^2$. As the present study area is not too large, we took as covariates X the grid cell heights provided by the WRF topography (see Fig. 2, right). Therefore, the term $X(\mathbf{s}) \cdot \boldsymbol{\alpha}_\kappa$ can be written as $\alpha_{\kappa 0} + h_s \alpha_{\kappa 1}$, so that we need to provide priors for $\alpha_{\kappa 0}, \alpha_{\kappa 1}$. Gaussian distributions were used for both cases:

$$\alpha_{\kappa j} \sim N(a_{\alpha_{\kappa j}}, b_{\alpha_{\kappa j}}^2), \qquad \kappa = \{\mu, \sigma, \xi\}, j = \{0, 1\},$$

where the hyperparameters of mean ($a_{\alpha_{\kappa j}}$) and variance ($b_{\alpha_{\kappa j}}^2$) were chosen appropriately in such a way that the distribution was either non- or only weakly informative with no extra information about the parameters, e.g., $\alpha_{\mu j} \sim N(0, 10000)$. A similar model (Gaussian) was selected for the trend parameter $\boldsymbol{\gamma}$ in the non-stationary model. The sill parameters $\beta_{\kappa 1}^2$ and the nugget variances τ_κ^2 were parametrized by inverse gamma distributions. A uniform distribution was chosen for the range $\beta_{\kappa 2}$.

3.4 Parameter estimation

Pulling together the different parts of the hierarchical model, and taking into account Bayes' theorem as expressed in Equation (1), for the stationary model, the posterior distribution of the parameters is given by the expression

$$\mathbf{P}\big(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \mathbf{W}_\mu, \mathbf{W}_\sigma, \mathbf{W}_\xi, \boldsymbol{\alpha}_\mu, \boldsymbol{\alpha}_\sigma, \boldsymbol{\alpha}_\xi, \boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\sigma, \boldsymbol{\beta}_\xi, \tau_\mu^2, \tau_\sigma^2, \tau_\xi^2 \mid \boldsymbol{Y}, \boldsymbol{X}\big)$$

$$\propto \prod_{s=1}^{ngrid} \prod_{t=1}^{ndat} \mathbf{GEV}(y_{ts}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi})$$

$$\cdot \mathbf{P}\big(\boldsymbol{\mu}|X_\mu, \boldsymbol{\alpha}_\mu, W_\mu, \tau_\mu^2\big)\mathbf{P}\big(\boldsymbol{\sigma}|X_\sigma, \boldsymbol{\alpha}_\sigma, W_\sigma, \tau_\sigma^2\big)\mathbf{P}\big(\boldsymbol{\xi}|X_\xi, \boldsymbol{\alpha}_\xi, W_\xi, \tau_\xi^2\big)$$

$$\cdot \mathbf{P}\big(\boldsymbol{\Psi}_\mu|\boldsymbol{\beta}_\mu\big)\mathbf{P}\big(\boldsymbol{\Psi}_\sigma|\boldsymbol{\beta}_\sigma\big)\mathbf{P}\big(\boldsymbol{\Psi}_\xi|\boldsymbol{\beta}_\xi\big)$$

$$\cdot \mathbf{P}(\alpha_{\mu 0})\mathbf{P}(\alpha_{\mu 1})\mathbf{P}(\alpha_{\sigma 0})\mathbf{P}(\alpha_{\sigma 1})\mathbf{P}(\alpha_{\xi 0})\mathbf{P}(\alpha_{\xi 1})$$

$$\cdot \mathbf{P}(\beta_{\mu 0}^2)\mathbf{P}(\beta_{\mu 1})\mathbf{P}(\beta_{\sigma 0}^2)\mathbf{P}(\beta_{\sigma 1})\mathbf{P}(\beta_{\xi 0}^2)\mathbf{P}(\beta_{\xi 1})\mathbf{P}(\tau_\mu^2)\mathbf{P}(\tau_\sigma^2)\mathbf{P}(\tau_\xi^2), \tag{12}$$

where $W_\kappa, \kappa = \mu, \sigma, \xi$ has been estimated by its definition $W_\kappa = Z\Psi_\kappa$. For the non-stationary model, the posterior distribution is given by the expression

$$\mathbf{P}\big(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \mathbf{W}_\mu, \mathbf{W}'_\mu, \mathbf{W}_\sigma, \mathbf{W}_\xi, \boldsymbol{\alpha}_\mu, \boldsymbol{\gamma}_\mu, \boldsymbol{\alpha}_\sigma, \boldsymbol{\alpha}_\xi,$$
$$\boldsymbol{\beta}_\mu, \boldsymbol{\beta}'_\mu \boldsymbol{\beta}_\sigma, \boldsymbol{\beta}_\xi, \tau_{\mu^2}, \tau'_{\mu^2}, \tau_\sigma^2, \tau_\xi^2 \mid \boldsymbol{Y}, \boldsymbol{X}, f(t)\big)$$

$$\propto \prod_{s=1}^{ngrid} \prod_{t=1}^{ndat} \mathbf{GEV}(y_{ts}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}, f(t))$$

$$\cdot \quad \mathbf{P}\big(\boldsymbol{\mu_1}|X_\mu, \boldsymbol{\alpha}_\mu, W_\mu, \tau_\mu^2\big)\mathbf{P}\big(\boldsymbol{\mu_2}|X'_\mu, \boldsymbol{\gamma}_\mu, W'_\mu, \tau'^2_\mu\big)$$

$$\cdot \quad \mathbf{P}\big(\boldsymbol{\sigma}|X_\sigma, \boldsymbol{\alpha}_\sigma, W_\sigma, \tau_\sigma^2\big)\mathbf{P}\big(\boldsymbol{\xi}|X_\xi, \boldsymbol{\alpha}_\xi, W_\xi, \tau_\xi^2\big)$$

$$\cdot \quad \mathbf{P}\big(\boldsymbol{\Psi}_\mu|\boldsymbol{\beta}_\mu\big)\mathbf{P}\big(\boldsymbol{\Psi}'_\mu|\boldsymbol{\beta}'_\mu\big)$$

$$\cdot \quad \mathbf{P}\big(\boldsymbol{\Psi}_\sigma|\boldsymbol{\beta}_\sigma\big)\mathbf{P}\big(\boldsymbol{\Psi}_\xi|\boldsymbol{\beta}_\xi\big)$$

$$\cdot \quad \mathbf{P}(\alpha_{\mu 0})\mathbf{P}(\alpha_{\mu 1})\mathbf{P}(\gamma_{\mu 0})\mathbf{P}(\gamma_{\mu 1})\mathbf{P}(\alpha_{\sigma 0})\mathbf{P}(\alpha_{\sigma 1})\mathbf{P}(\alpha_{\xi 0})\mathbf{P}(\alpha_{\xi 1})$$

$$\cdot \quad \mathbf{P}(\beta_{\mu 0^2})\mathbf{P}(\beta_{\mu 1})\mathbf{P}(\beta'_{\mu 0^2})\mathbf{P}(\beta'_{\mu 1})\mathbf{P}p(\beta_{\sigma 0}^2)\mathbf{P}(\beta_{\sigma 1})\mathbf{P}(\beta_{\xi 0}^2)\mathbf{P}(\beta_{\xi 1})$$

$$\cdot \quad \mathbf{P}(\tau_\mu^2)\mathbf{P}(\tau'^2_\mu)\mathbf{P}(\tau_\sigma^2)\mathbf{P}(\tau_\xi^2), \tag{13}$$

where $\boldsymbol{\mu} = \boldsymbol{\mu}_0 + f(t)\boldsymbol{\mu}_1$, with $f(t) = (t - t_0)$ being the linear temporal-trend function.

The simulation of the posterior distribution was carried out by means of an MCMC method, in particular using a Gibbs sampler with embedded Metropolis-Hastings steps

(see Gilks et al. (1996) for more details about MCMC methods). The MCMC was iterated for 50 000 samples with a burn-in period of 30 000 samples to allow the MCMC to reach the stationary state. To avoid the autocorrelation that appears in the MCMC, we kept one out of ten samples for the subsequent calculations. To test the convergence of the chains, four chains were constructed starting with different values of the parameters being simulated. The Gelman–Rubin diagnostic convergence test (see Cowles and Carlin (1996)) was used to evaluate the convergence of these four chains. In most cases, that burn-in period was sufficient for convergence to be reached.

A preliminary study of the spatial models revealed that the spatial-trend coefficient $\boldsymbol{\alpha}_\xi$ for the shape parameter ξ is not significantly different from zero. For this reason, and given that the shape parameter is the most difficult to estimate Sang and Gelfand (2009); Cooley and Sain (2010) we shall assume this shape parameter to be constant throughout the domain.

The code used to carry out the simulations was written in FORTRAN, following quite closely the procedure in Finley et al. (2015). For the Gelman-Rubin diagnostic test, the CODA package of the R language was used, and the maps and figures were prepared using the R packages *fields* and *sp*.

3.5 Checking the models

An important step in a statistical analysis is to assess whether the observed data are indeed fitted by the proposed statistical model or models. A two-step procedure was followed. In the first step, the models were ranked according to the WAIC (Widely Applicable Information Criterion) model comparison tool (Gelman et al. 2014). In the second step, the chosen model was contrasted with the data by means of a Bayesian p-value procedure. In both steps the posterior predictive distribution was used. The posterior predictive distribution (PPD) is defined as (Gelman et al. 1996)

$$\mathbf{P}(y^{\text{rep}}|M, y) = \int \mathbf{P}(y^{\text{rep}}|M, \theta)\mathbf{P}(\theta|M, y)d\theta, \tag{14}$$

which gives the probability of obtaining new replicated data given the model M and the observed data y. In this equation, θ represents the parameters of the model, $\mathbf{P}(\theta|M, y)$ the posterior distribution, and $\mathbf{P}(y^{\text{rep}}|M, \theta)$ the likelihood.

The WAIC criterion to rank the models is defined by the equation

$$\text{WAIC} = -2\log\left(\prod_{i=1}^{n} \mathbf{P}(y_i)\right) + p_{\text{waic}} = -2\sum_{i=1}^{n}\log\left(\mathbf{P}(y_i)\right) + p_{\text{waic}}, \tag{15}$$

where $\mathbf{P}(y_i)$ is the PPD given by Equation (14), and p_{waic} is a term that measures both the bias introduced into the test from using the data twice—once to estimate the parameters, and a second time to use them in the test (see Gelman et al. (2014))—and the complexity of the model (penalizing models with more parameters). The term $-2\sum_{i=1}^{n}\log\left(\mathbf{P}(y_i)\right)$ is estimated by the equation

$$-2\sum_{i=1}^{n}\log\left(\mathbf{P}(y_i)\right) \approx -2\sum_{i=1}^{n}\log\left[\frac{1}{J}\sum_{j=1}^{J}\left(\mathbf{P}(y_i|\theta^j)\right)\right] \tag{16}$$

where J is the number of samples taken from the MCMC method and θ^j are the estimated parameters of the model. The p_{waic} parameter is defined by the expression

$$p_{\text{waic}} = \sum_{i=1}^{n} \text{var}_{\text{post}}\left[\log\left(\mathbf{P}(y_i)\right)\right] \tag{17}$$

where var_{post} is the operator variance and is estimated as

$$\text{var}_{\text{post}}(x) = \frac{1}{J-1}\sum_{j=1}^{J}(x_j - \bar{x}); \qquad x_j = \log\left(\mathbf{P}(y_i|\theta^j)\right). \tag{18}$$

As before, J is the number of samples taken from the MCMC. The choice of the WAIC criterion is motivated by the fact that it is easy to calculate from the MCMC and that it has the property of being asymptotically equivalent to a leave-one-out cross-validation (LOO-CV) test (see Gelman et al. (2014)).

In the second step, the Bayesian p-value was evaluated as follows. Let $T(y)$ be some feature of the data we are interested in, for example, the skewness of the data, and we want to see whether the model predicts this feature reasonably well. A measure of the agreement is the Bayesian p-value defined by Lynch and Bruce (2004)

$$p_b(y) = \mathbf{P}\left(T(y^{\text{rep}}) \geq T(y) \mid M, y\right). \tag{19}$$

A p_b-value that is too small or too large would indicate that the model does not reproduce the data (or at least some features of the data) well. One way to calculate the p_b-value is from the MCMC. As was noted above, an MCMC method allows one to obtain values of the parameter θ from the posterior distribution $\mathbf{P}(\theta|y)$. For each simulated θ value, a y^{rep} value is simulated by using the posited model, and then the statistic $T(y^{\text{rep}})$ is computed. Evaluating the number of times that $T(y^{\text{rep}}) \geq T(y)$ gives the p_b-value as

$$p_b = \frac{\#\left(T(y^{\text{rep}}) \geq T(y)\right)}{J}, \tag{20}$$

with J being the number of samples of θ taken from the Markov chain.

3.6 Prediction at a non-grid site

Once we have fitted the model, we are in a position to predict variables of interest at a non-grid site. In particular, we are interested in predicting the parameters of the GEV distribution $\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}$ in order to estimate, for example, the T-year return levels at those non-grid sites. We shall illustrate the method developed for the location parameter $\boldsymbol{\mu}$. Based on Equation (3), we shall assume that if μ_0 is the location parameter at a non-grid site then the joint statistical distribution of the location parameter at the grid and non-grid sites is given by

$$\begin{pmatrix} \mu_0 \\ \boldsymbol{\mu} \end{pmatrix} \sim \mathbf{MVN} \begin{pmatrix} m_0 & | & \sigma_0 & \boldsymbol{\Sigma}_{0n} \\ \mathbf{m} & | & \boldsymbol{\Sigma}_{n0} & \boldsymbol{\Sigma}_{nn} \end{pmatrix} \tag{21}$$

where **MVN** represents the multivariate normal distribution, $m_0, \mathbf{m}$ are the means of the distributions at the non-grid site and grid sites respectively, σ_0 is the variance at the non-grid site, $\boldsymbol{\Sigma}_{nn}$ is the $n \times n$ covariance matrix at the grid sites, and $\boldsymbol{\Sigma}_{0n}, \boldsymbol{\Sigma}_{n0}$ are the $1 \times n, n \times 1$ covariance matrices between the non-grid and grid sites. A result from multinormal distribution theory allows us to say that the conditional distribution of μ_0 given $\boldsymbol{\mu}$, i.e., $P(\mu_0|\boldsymbol{\mu})$, is normal with mean given by

$$m_1 = m_0 + \boldsymbol{\Sigma}_{0n}\boldsymbol{\Sigma}_{nn}^{-1}(\boldsymbol{\mu} - \mathbf{m}) \tag{22}$$

and variance

$$\sigma_1 = \sigma_0 - \boldsymbol{\Sigma}_{0n}\boldsymbol{\Sigma}_{nn}^{-1}\boldsymbol{\Sigma}_{n0}. \tag{23}$$

Once the mean m_1 and variance σ_1 have been estimated, one can take samples of the location parameter at the non-grid site. According to Eqs. (3) to (5), the $\boldsymbol{\mu}$ distribution is

$$\boldsymbol{\mu} \sim \mathbf{MVN}(X \cdot \boldsymbol{\alpha}_u + W_\mu, \tau_\mu^2 \mathbf{I}) \tag{24}$$

where W_μ is the spatial process given by $W_\mu = Z\Psi_\mu$, with Z being an $n \times r$ matrix and Ψ_μ an $r \times r$ MVN process with

zero mean and variance $\boldsymbol{\Sigma}_\mu$ (Eqs. (8) and (9)). Integrating in Ψ_μ, the above distribution may be expressed as

$$\boldsymbol{\mu} \sim \mathbf{MVN}(X \cdot \boldsymbol{\alpha}_\mu, Z\boldsymbol{\Sigma}_\mu Z^T + \tau_\mu^2 \mathbf{I}). \tag{25}$$

From this expression, one can take $\mathbf{m} = X \cdot \boldsymbol{\alpha}_\mu$ and $m_0 = X_0 \cdot \boldsymbol{\alpha}_\mu$. Taking into account Eq. (10), the $r \times r$ covariance matrix $\boldsymbol{\Sigma}_\mu$ is given by

$$\boldsymbol{\Sigma}_\mu(\mathbf{s}', \mathbf{s}') = \beta_{\mu 1}^2 \exp(-\|\mathbf{s}_i' - \mathbf{s}_j'\|)/\beta_{\mu 2}), \qquad i,j = 1, \cdots, r. \tag{26}$$

The Z matrix is given by (see Eq. (11)):

$$Z = \boldsymbol{\Sigma}_\mu(\mathbf{s}, \mathbf{s}') \cdot \boldsymbol{\Sigma}_\mu^{-1}(\mathbf{s}', \mathbf{s}'). \tag{27}$$

Lastly, the covariance between the non-grid and a grid site is taken as

$$\boldsymbol{\Sigma}_{n0} = \beta_{\mu 1}^2 \exp(-||\mathbf{s}_i - \mathbf{s}_0||)/\beta_{\mu 2}), \tag{28}$$

and $\sigma_0 = \tau_\mu^2$.

4 Results

First of all, it is important to know whether the climate of the region (represented by the summer maximum temperature) is well described by the WRF model. Figure 4 shows a map of the mean summer extreme temperatures for the study period (1981–2015) given by the Stead database Serrano-Notivoli et al. (2019) (left) and by the WRF model (right). As can be seen, the spatial distribution of the maximum temperature is quite similar in the two maps, so that we have reasonable confidence in the results given by the WRF model.

Moving to the statistical model, one first needs to establish the external covariate term X. Because the spatial domain is not too large, and given the influence of altitude

Fig. 4 Mean of the summer extreme temperature during the period 1981–2015 given by the WRF model (right) for the Iberian Peninsula and during the period 1981–2014 given by Stead database for Spain. Scales are in °C
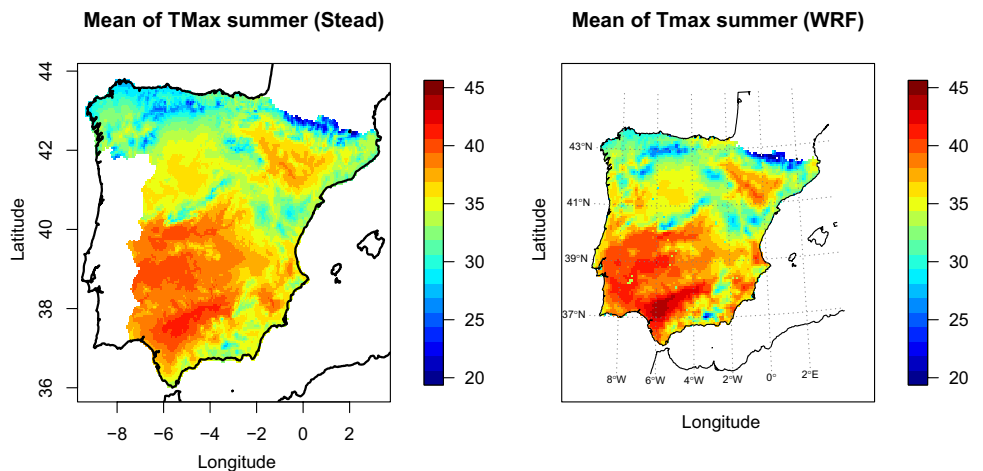
on temperature, we chose the height of the grid cell to be our covariate. Prior to their introduction into the model, the heights were linearly normalized to the interval (0, 1). With this choice of covariate, the spatial-trend term in the location parameter $\boldsymbol{\mu}$ (stationary models) and $\boldsymbol{\mu}_0$ (non-stationary models) is

$$\mathbf{X}_\mu(s) \cdot \boldsymbol{\alpha}_\mu = \alpha_0 + h_s \alpha_1, \tag{29}$$

and in the temporal-trend coefficient $\boldsymbol{\mu}_1$ is

$$\mathbf{X}'_\mu(s) \cdot \boldsymbol{\gamma}_\mu = \gamma_0 + h_s \gamma_1. \tag{30}$$

Prior to its use in the model, the linear temporal-trend function $f(t) = t - t_0$ was also linearly normalized to the interval $(-1, 1)$.

We posited various models depending on whether or not there exists a temporal trend in the location parameter, and, within them, whether there exists a dependence on height in both the location and the scale parameters. As mentioned above, to rank the proposed models, the WAIC criterion was used. The results are presented in Table 2.

Observing column p_{waic}, one sees that as the complexity of the models increases so does the complexity parameter. The results in the last column lead to non-stationary models fitting better than stationary models, with model 2110 performing the best. Therefore, this model was chosen for the next step—the assessment of the model against the data using the Bayesian p-value.

As noted above, to apply the Bayesian p-value we need to choose a statistic $T(y)$ of the data that can help determine which model best explains some feature of the data. Because we are considering a time dependence in the models, we

took the Mann-Kendall test as the feature to explain. This is a non-parametric test used to determine whether a uniform trend exists in a set of independent data. The test statistic is given by the expression Gallego et al. (2011)

$$T(y) = \frac{1}{\sqrt{n(n-1)(2n+5)/18}} \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sgn}(y_j - y_i) - \text{sgn}(T) \right), \tag{31}$$

where $\text{sgn}(x)$ is the sign function defined as $+1$ if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$, and n is the number of data for each grid cell. We calculated the p_b-value by evaluating this statistic for the 'observed' and the replicated data. To simulate replicated data according to the PPD given by expression (14), an instance of the μ, σ, ξ parameters is taken from the MCMC chain, and a sample of n elements is extracted using the GEV distribution. The statistic $T(y^{\text{rep}})$ is then evaluated from this sample. This procedure is repeated for each element of the chain, obtaining a sample of J replicated Mann-Kendall statistics. This statistic is also evaluated for the observed data. From the replicated sample and the observed datum, the p_b value is calculated as the percentile of the observed $T(y^{\text{obs}})$ in the J replicated Mann-Kendall statistics. Values of p_b that are too large or too small mean that the replicated data do not explain that feature of the observed data well. If there are N grid cell elements in the domain, we obtain N p_b values. From this set, we calculated, by way of synthesis, the minimum, the 25, 50, and 75 percentiles, and the maximum. For the chosen 2110 model, these values were 0.1870, 0.4080, 0.4470, 0.5440, 0.9500, respectively. Values of the 25, 50, and 75 percentiles are close to 0.5, which means that for most sites the 2110 model explains quite well the temporal trend observed in the results given by the RCM.

Let us now consider this model's results in greater detail. Firstly, Fig. 5 shows the spatial regression coefficients α_0, α_1. The right-hand panel shows that the height-trend coefficient is negative with a mean value of $-15.1°$C. Therefore, given that the heights were normalized, and considering that the maximum altitude in the model is 1794.8 m and the minimum is 134.8 m, one has the result that $\alpha_1 = -9.35$ °C/km, which is quite close to the theoretical adiabatic coefficient of -9.8 °C/km. It is important to bear in mind that the data correspond to summer when the temperatures are the highest and the atmospheric boundary layer is in nearly adiabatic equilibrium. This result lends added confidence to the proposed model.

Secondly, we shall show the results for the temporal-trend coefficients γ_0, γ_1. Although the chosen model only has the γ_0 coefficient, an analysis of that model with γ_1 present showed that coefficient to not differ from zero, i.e., the temporal-trend coefficient has no height dependence.

Table 2 Summary of the values of the WAIC information criterion

Kind	Model	p_{waic}	WAIC
Stationary	200	816.7	232,663
	210	923.8	231,042
	220	923.4	231,038
Non-stationary	2000	819.9	231,314
	2010	1051.2	230,790
	2020	1050.1	230,782
	2110	1079.4	230,480
	2220	1079.1	230,486

The column headed 'Model' lists the tags used to distinguish the spatial-trend parameters used in each statistical model. These tags are: 2, dependence on the observatory's altitude; 1, no dependence on altitude; and 0, the parameter is considered constant throughout the domain. For the stationary models, the three parameters are location, scale, and shape, respectively. For the non-stationary models, the four parameters are location $\boldsymbol{\mu}_0$, $\boldsymbol{\mu}_1$, and scale and shape, respectively. Column p_{waic} lists the model complexity parameter, and column WAIC, the WAIC parameter
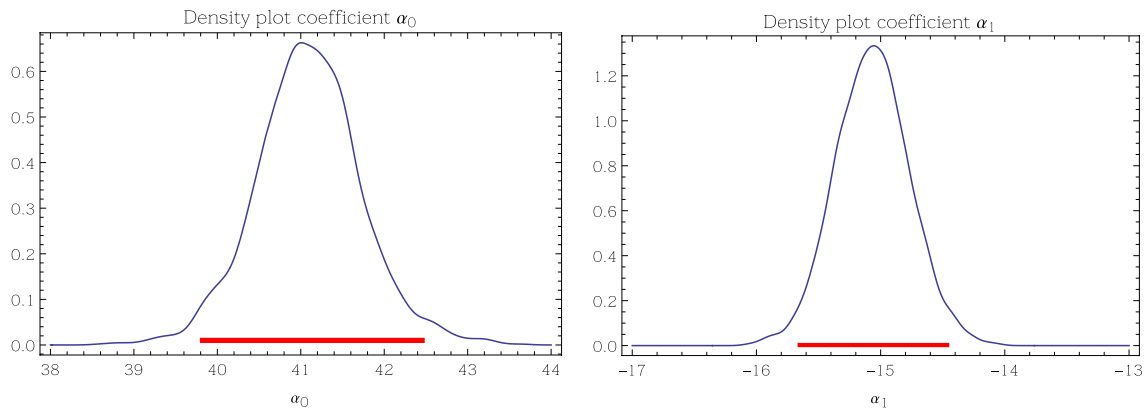
Fig. 5 Spatial-regression coefficients α_0 (left) and α_1 (right). The red line shows the interval (2.5%, 97.5%)
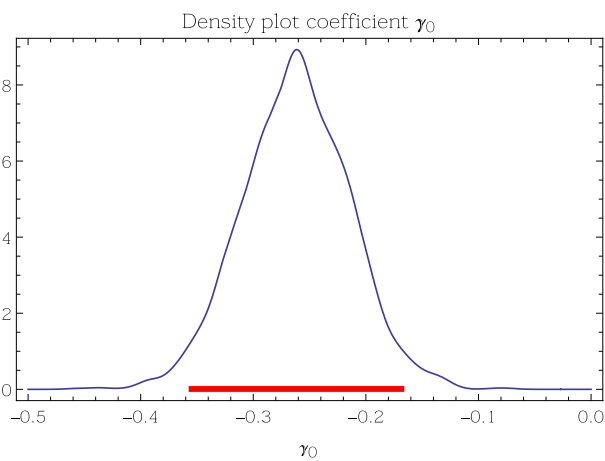

Fig. 6 Temporal-trend coefficients γ_0. The red line shows the interval (2.5%, 97.5%)

Figure 6 shows a density plot of γ_0. The mean (2.5%, 97.5%) is −0.26 (−0.39, −0.12). From the aforementioned linear temporal-trend function, one finds a temporal trend of − 0.15 °C per decade with a 2.5–97.5% interval of (−0.21 °C, −0.10 °C) per decade. This is a surprising result because

one would expect an increase in temperature due to climate change. Figure 7 shows a plot of the temporal-trend regression coefficient (°C/year) obtained by the Sen method (Sen 1968; Hirsch and Smith 1982) for the summer extreme 2-metre-height temperature in the ERA Interim reanalysis used to externally force the WRF model in the study period 1981–2015. The ERA Interim reanalysis shows a predominantly positive trend for the study period (1981–2015), mainly in the central area of the Iberian Peninsula. However, this positive trend tends to decrease towards the west of the IP, with some regions showing a negative, although small, trend. Also, one can see from the figure that the WRF model shows a negative trend throughout the west, including the study region, and a positive trend in the center of the IP. It is possible, however, that internal variability and model variability in the WRF model which persist after the constraint of the boundary conditions given by the external reanalysis (Alexandru et al. 2009; Hawkins and Sutton 2009) lead to the enhanced, although small, negative trends in extreme temperatures appearing in a simulation as short as that of the present work (35 years).

Some previous studies on temporal trends in RCM simulations have shown the difficulty these models have

Fig. 7 Temporal-trend regression coefficient (°C/year) obtained by the Sen method for summer extreme 2-metre-height temperature in the ERA Interim reanalysis (left) and in the WRF model (right)
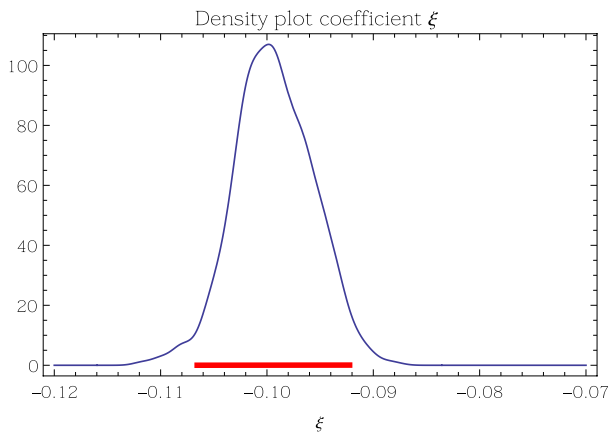
Density plot coefficient ξ
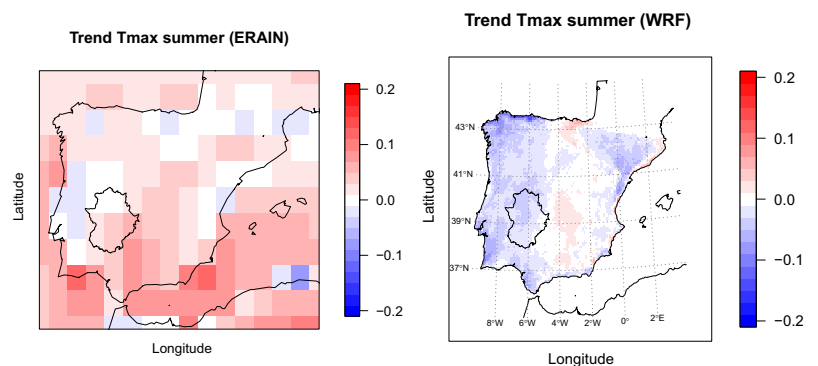
Fig. 8 Shape parameter ξ. The red line shows the interval (2.5%, 97.5%)

in estimating these trends. In Lorenz and Jacob (2010), temporal trends in 2-m seasonal and annual average temperatures were simulated by 13 RCMs driven by ERA40 reanalysis for the period 1960–2000 within the framework of the ENSEMBLES EU-Project. The trends of the annual average temperature obtained with the RCMs were found to underestimate the observed trends in all regions, and in most regions they also underestimated the trends given by the ERA40 reanalysis dataset. Bukovsky (2012) analysed temporal trends in 2-m seasonal average temperatures obtained from six RCMs driven by NCEP-2 reanalysis for the period 1980–2003 within the framework of the North American Regional Climate Change Assessment Program (NARC-CAP). According to those authors' own words: 'There are no clear conclusions about the behaviour of the RCMs because some of the models cannot capture certain seasonal trends over portions of the domain' Min et al. (2013).

An interesting parameter to study is the shape parameter, which provides information on the shape of the extreme temperature probability distribution function (PDF). A positive value indicates that there is no upper bound on the extreme temperatures, and a negative value that there is an upper bound. Figure 8 shows the posterior distribution of this parameter for the chosen model.

The mean (2.5%, 97.5%) is −0.994 (−0.092, −0.106), thus supporting a bound on the extreme temperatures for the region under study. The negative value of the shape parameter in the GEV and Generalized Pareto Distribution used to fit the extreme temperature distribution seems to be a universal feature because several other works have found similar negative values for different parts of the world (see, for example, Salleh and Hasan (2018) for Malaysia, García-Cueto et al. (2013) for Mexico, Furrier et al. (2010) for USA, Brown et al. (2008) for different parts of the world, Nogaj

et al. (2006) for the North Atlantic, and Parey et al. (2007) for France).

It is important to note that we also performed the model's calculations with other subsampling values (2×2, 4×4), obtaining results similar to those just shown for the 3×3 subsampling case.

Once the parameters of the GEV distribution are known, one could already display maps of them. From the perspective of field practitioners, however, it would be more interesting to combine them and display return level maps. The T-year return level is the quantile for which the probability that the annual maximum exceeds this quantile is $1/T$ (see Kharin and Zwiers (2005) and Cooley (2013) for a definition of the T-year return level and T-year return period in a non-stationary context). From the GEV distribution, one gets the expression

$$y_p(s) = \mu_s - \frac{\sigma_s}{\xi_s}\left(1 - (-\log(1-p))^{-\xi_s}\right), \tag{32}$$

with $p = 1/T$. For the non-stationary model (Equation (7)), the T-year return level is

$$y_p(s,t) = \mu_{0s} + \mu_{1s} \cdot f(t) - \frac{\sigma_s}{\xi_s}\left(1 - (-\log(1-p))^{-\xi_s}\right). \tag{33}$$

From this equation, one can evaluate the difference between the end and the beginning of the period,

$$\Delta y_p(s) = 2\mu_{1s}, \tag{34}$$

considering the normalization used in $f(t)$.

Figure 9 (left) shows a map of the 40-year (97.5 percentile) return level. As may be appreciated by comparing this figure with that of the topography (Fig. 2), the spatial distribution of the 40-year return level is completely determined by the topography as should be expected for a variable such as temperature. We also calculated the difference in the 40-year return level between the end and the beginning of the study period. The resulting map is shown in Fig. 9 (right). One sees that, for most of the region, there is a decrease in the return level by about 0.8 °C to 1.0 °C, except at the eastern edge of the region where the decrease approaches 0 °C

One of the advantages of having a spatial model such as that used in this work is the possibility of predicting the GEV distribution parameters at a non-observed site. In this sense, we evaluated the T-year return period for a set of 28 weather stations in the region under study. To obtain values of the location and scale parameters, we drew from a normal distribution with mean and variance given by Equations (22) and (23), and with Equation (32) we got the T-year return period. Figure 10 is a scatter plot of the predicted versus observed 40-year return levels, together with the regression line through

Fig. 9 Mean of the 40-year return level for t=0 (left), and $\Delta y_{0.05}$ (right). Scales are in °C
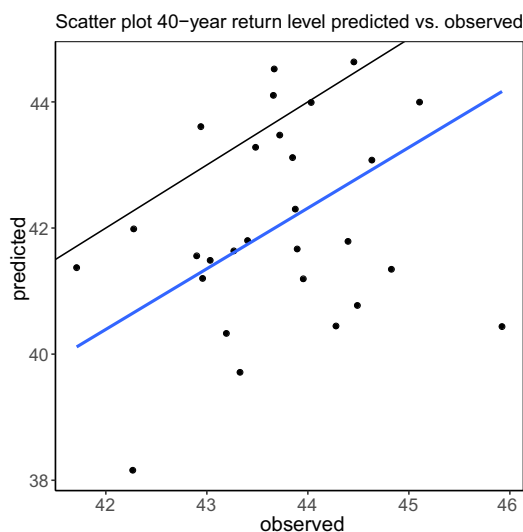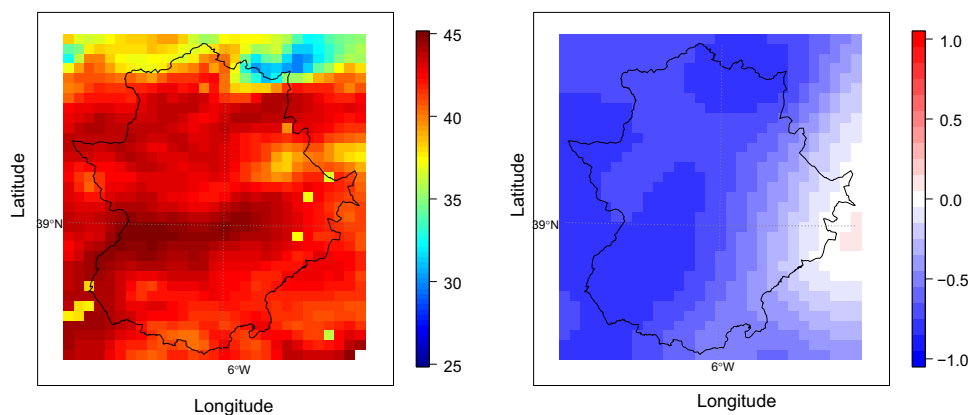

Fig. 10 Scatter plot of the predicted and 'observed' 40-year return levels for 28 observatories in the region. The black line is the 1:1 line, and the blue line the $y \sim x$ linear regression. Scales are in °C

the origin $y \sim x$ (blue) and the 1:1 line (black). The observed 40-year return level was obtained by fitting a GEV model to each observed data set using the R package *ismev* (Coles (2001)). The regression coefficient is 0.96 ($\pm$0.014), the bias (observed - predicted) is 1.66 °C, and the RMS is 2.30°C. The results seem quite promising, although the predicted values are underestimates of those observed.

5 Summary and conclusions

In the present work, we have described a two-step study of extreme temperatures in the region of Extremadura (Spain). In the first step, we used the WRF model to obtain a set of extreme temperatures for the period 1981–2015 in our region. The boundary conditions needed by the model were obtained from ECMWF's ERA Interim data. In the second

step, a statistical study was made of the extreme temperature data obtained in the region. To this end, a Bayesian hierarchical spatio-temporal model with a GEV parametrization of the extreme event data was used. Two advantages of this kind of model are that it reduces the uncertainty in the parameters of the statistical distribution by pooling data among 'observatories' in a consistent way, and that it can project important features of the statistical distribution, such as the T-year return level, to non-observing sites.

The spatial-trend parameter given by the model chosen was quite consistent with that corresponding to the dry adiabatic gradient, i.e., the gradient one would expect in the summer season when adiabatic conditions are prevalent in the atmospheric boundary layer. The negative trend in the location parameter is consistent with the negative trend that the WRF model shows for the extreme temperature in the western part of the Iberian Peninsula which has enhanced the trend given by ERA Interim for the same region. A possible reason for this result may lie in the internal and model uncertainties in using the WRF model. The shape parameter of the GEV model was negative, showing that there is an upper bound to the extreme temperatures over the region.

An important product we have obtained from the statistical model is that of obtaining return levels at non-observing sites. A comparison of the 40-year return levels of observed temperatures with those predicted by the statistical model showed a bias (observed minus predicted) of 1.66°C.

Declarations

Conflict of interest The authors have not disclosed any competing interests.

Code availability Enquires about the code availability should be directed to the authors.

References

Alexandru A, Elia R, Laprise R, Separovic L, Biner S (2009) Sensitivity study of regional climate model simulations to large-scale nudging parameters. Mon Weather Rev 137:1666–1686. https://doi.org/10.1175/2008MWR2620.1

Banerjee S, Gelfand AE, Finley AO, Sang H (2008) Gaussian predictive process models for large spatial data sets. J. Royal Statistical Society, B 70(4):825–848

Barriopedro D, Fischer EM, Luterbacher J, Trigo RM, García-Herrera R (2011) The hot summer of 2010: redrawing the temperature record map of Europe. Science 332:220–224. https://doi.org/10.1126/science.1201224

Bartolomeu S, Carvalho MJ, Marta-Almeida M, Melo-Gonçalves P, Rocha A (2016) Recent trends of extreme precipitation indices in the Iberian Peninsula using observations and WRF model results. Physics and Chemistry of the Earth, Parts A/B/C 94:10–21. https://doi.org/10.1016/j.pce.2016.06.005. 3rd International Conference on Ecohydrology, Soil and Climate Change, EcoHCC'14

Berliner LM (2003) Physical-statistical modeling in geophysics. Journal of Geophysical Research: Atmospheres 108(D24). https://doi.org/10.1029/2002JD002865

Brown SJ, Caesar J, Ferro CAT (2008) Global changes in extreme daily temperature since 1950. J. Geophysical Reseach 113:05115. https://doi.org/10.1029/2006JD008091

Bukovsky MS (2012) Temperature trend in the NARCACAP regional climate models. J Clim 25:3985–3991. https://doi.org/10.1175/JCLI-D-11-00588.1

Casson E, Coles S (1999) Spatial Regression Models for Extremes. Extremes 1(4):449–468

Chen F, Dudhia J (2001) Coupling an advanced land surface-hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. Monthly Weather Review 129: 569–585. https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2

Coles S (2001) An Introduction to Statistical Modeling of Extreme Values. Springer, London, p 208

Cooley D (2013) Return periods and return levels under climate change. In: AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., Sorooshian, S. (eds.) Extremes in a Changing Climate, pp. 97–114. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-4479-0

Cooley D, Sain SR (2010) Spatial Hierarchical Modeling of Precipitation Extremes From a Regional Climate Model. J Agric Biol Environ Stat 15(3):381–402

Cooley D, Nychka D, Naveau P (2007) Bayesian spatial modeling of extreme precipitation return levels. J Am Stat Assoc 102(479):824–840

Cowles MK, Carlin BP (1996) Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. J Am Stat Assoc 91(434):883–904

Craigmile PF, Guttorp P (2013) Can regional climate model reproduce observed extreme temperature? Statistica (Bologna) 73:103–122

Cressie N, Wikle CK (2011) Statistics for Spatio-Temporal Data. Wiley, Amsterdam

Cubasch U, Wuebbles D, Chen D, Facchini MC, Frame D, Mahowald N, Winther J-G (2013) Introduction. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex v, Midgley PM (eds.) Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK and NY USA

Davison AC, Padoan SA, Ribatet M (2012) Statistical modeling of spatial extremes. Stat Sci 27(2):161–186. https://doi.org/10.1214/11-STS376

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Holm EV, Isaksen L, Kallberg P, Kohler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597. https://doi.org/10.1002/qj.828

Deng Z, Xin Q, Liu J, Madras N, Wang X, Zhu H (2016) Trend in frequency of extreme precipitation events over Ontario from ensembles of multiple GCMs. Clim Dyn 46:2909–2921. https://doi.org/10.1007/s00382-015-2740-9

Eidsvik J, Finley AO, Banerjee S, Rue H (2012) Approximate bayesian inference for large spatial datasets using predictive process models. Computationl Statistics and Data Analysis 56:1362–1380

Epstein ES (1985) Statistical Inference and Prediction in Climatology: A Bayesian Approach. Meteorological Monographs, vol. 20. American Meteorological Society, New York

Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, Mastrandrea MD, Mach KJ, Plattner G-K, Allen SK, Tignor M, Midgley PMe (2012) Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Cambridge University Press, Cambridge, UK and NY USA

Finley AO, Sang H, Banerjee S, Gelfand AE (2009) Improving the performance of predictive process modeling for large datasets. Computationl Statistics and Data Analysis 53:2873–2884

Finley AO, Banerjee S, Gelfand AE (2015) spBayes for Large Univariate and Multivariate Point-Referenced Spatio-Temporal Data Models. J Stat Softw 63(13):1–28

Fita L, Polcher J, Giannaros TM, Lorenz T, Milovac J, Sofiadis G, Katragkou E, Bastin S (2019) CORDEX-WRF v1.3: development of a module for the Weather Research and Forecasting (WRF) model to support the CORDEX community. Geoscientific Model Development 12(3):1029–1066. https://doi.org/10.5194/gmd-12-1029-2019

Fuentes M, Guttorp P, Challenor P (2003) Statistical Assessment of Numerical Models. Int Stat Rev 71(2):201–222

Furrier EM, Katz RW, D, WM, Furrer R (2010) Statistical modeling of hot spells and heat waves, Climate Res 43:191–205

Gallego MC, Trigo RM, Vaquero JM, Brunet M, García JA, Sigró J, Valente MA (2011) Trends in frequency indices of daily precipitation over the Iberian Peninsula during the last century. J. Geophysical Reseach 116:02109. https://doi.org/10.1029/2010JD014255

García-Cueto OR, Santillan-Soto N, Quintero-Muñoz M, Ojeda-Benitez S, Velázquez-Limon N (2013) Extreme temperature scenarios in mexicali, mexico under climate change conditions. Atmosfera 26:509–520

García-Herrera R, Díaz J, Trigo RM, Luterbacher J, Fischer EM (2010) A Review of the European Summer Heat Wave of 2003. Crit Rev Environ Sci Technol 40(4):267–306. https://doi.org/10.1080/10643380802238137

Gelfand AE, Zhu L, Bradley PC (2001) On the Change of Support Problem for Spatio-Temporal Data. Biostatistics 2(1):31–45

Gelman A, Meng X-L, Stern H (1996) Posterior predictive assessment of model fitness via realized discrepancies (with discussion). Stat Sin 6:733–807

Gelman A, Hwang J, Vehtari A (2014) Understanding predcitive information criteria for bayesian models. Stat Comput 24:997–1016. https://doi.org/10.1007/s11222-013-9416-2

Gilks WR, Richardson S, Spiegelhalter DJ (1996) Introducing Markov Chain Monte Carlo. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (eds.) Markov Chain Monte Carlo in Practice. Chapman & Hall

Grell GA, Freitas SR (2014) A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. Atmos Chem Phys 14:5233–5250. https://doi.org/10.5194/acp-14-5233-2014

Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. Bull Am Meteor Soc 90:1095–1107. https://doi.org/10.1175/2009BAMS2607.1

Hirsch RM, Smith RA (1982) Techniques of trend analysis for monthly water quality data. Water Resour Res 18:107–121

Hong S-Y (2010) A new stable boundary-layer mixing scheme and its impact on the simulated East Asian summer monsoon. Q.J.R. Meteorol. Soc 136:(1481-1496). https://doi.org/10.1002/qj.665

Hong S-Y, Lim JOJ (2006) The WRF Single moment 6-Class Microphysics Scheme (WSM6). Journal of the Korean Meteorological Society 42(2):129–151. https://doi.org/10.1175/MWR3199.1

Hong S-Y, Noh Y, Dudhia J (2006) A new vertical diffusion package with an explicit treatment of entrainment processes. Mon Weather Rev 139(9):2318–2341. https://doi.org/10.1175/MWR3199.1

Iacono MJ, Delamere JS, Mlawer EJ, Shephard MW, Clough SA, Collins WD (2008) Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. J Geophys Res 113:13103. https://doi.org/10.1029/2008JD009944

Jiang Z, Li W, Xu JJ, Li L (2015) Extreme precipittion indices over China in CMIP5 models. Part I: Model evalution. J. of Climate 28, 8603–8619. https://doi.org/10.1175/JCLI-D-15-0099.1

Jiménez PA, Dudhia J, González-Rouco JF, Navarro J, Montávez JP, García-Bustamente E (2012) A Revised Scheme for the WRF Surface Layer Formulation. Monthly Weather Review 140(3), 898–918. https://doi.org/10.1175/MWR-D-11-00056.1

Kharin VV, Zwiers FW (2005) Estimating extremes in transient climate change simulations. J Clim 18:1156–1173

Lorenz P, Jacob D (2010) Validationof temperature trends in the ENSEMBLES regional climate model runs driven by ERA40. Climate Res 44:167–177. https://doi.org/10.3354/cr00973

Lorenz R, Argüeso D, Donat MG, Pitman AJ, van den Hurk B, Berg A, Lawrence DM, Cheéruy F, Ducharne A, Hagemann S, Meier A, Milli PCD, Sereviratne SI (2016) Influence of land-atmosphere feedbacks on temperature and precipitation extremes in the GLACE-CMIP5 ensemble. Journal of Geophysical Research: Atmospheres 121:607–623. https://doi.org/10.1002/2015JD024053

Lynch SM, Bruce W (2004) Bayesian posterior predictive checks for complex models. Sociological Methods & Research 32(3):301–335. https://doi.org/10.1177/0049124103257303

Min E, Hazeleger W, van Oldenborgh GJ, Sterl A (2013) Evaluation of trends in high temperature extremes in north-western Europe in regional climate models. Environ Res Lett 8:014011. https://doi.org/10.1088/1748-9326/8/1/014011

Nogaj M, Yiou P, Parey S, Malek F, Naveau P (2006) Amplitude and frequency of temperature extremes over the North Atlantic region. Geophysical Research Letters 33(L10801). https://doi.org/10.1029/2005GL024251

Parey S, Malek F, Laurent C, Dacunha-Castelle D (2007) Trends and climate evolution: Statistical approach for very high temperatures in france. Clim Change 81:331–352. https://doi.org/10.1007/s10584-006-9116-4

Peterson TC, Folland CK, Gruza G, Hogg WD, Mokssit A, Plummer N (2001) Report on the Activities of theWorking Group on Climate Change Detection and Related Rapporteurs. WCDMP-47, WMO-TD 1071, WMO

Peterson TC (2005) Climate change indices. Bulletin of the World Meteorological Organization 54:83–86

Randall DA, Wood RA, Bony S, Colman R, Fichefet T, Fyfe J, Kattsob V, Pitman A, Shukla J, Srinivasas J, Stouffer RJ, Sumi A, Taylor KE (2007) Climate Models and Their Evaluation. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Renard B (2011) A Bayesian hierarchical approach to regional frequency analysis. Water Resour Res 47:11513. https://doi.org/10.1029/2010WR010089

Salleh NHM, Hasan H (2018) Generalized Pareto Distribution for Extreme Temperatures in Peninsular Malaysia . Science International(Lahore) 30:63–67

Sang H, Gelfand AE (2009) Hierarchical modeling for extreme values observed over space and time. Environ Ecol Stat 16:407–426. https://doi.org/10.1007/s10651-007-0078-0

Schliep EM, Cooley D, Sain SR, Hoeting J (2010) A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. Extremes 13:219–239. https://doi.org/10.1007/s10687-009-0098-2

Sen PK (1968) Estimates of the regression coefficient based on kendall's tau. J Am Stat Assoc 63:1379–1389

Serrano-Notivoli R, Beguería S, de Luis M (2019) Stead: a high-resolution daily gridded temperature dataset for spain. Earth System Science Data 11(3):1171–1188. https://doi.org/10.5194/essd-11-1171-2019

Sillmann J, Kharin V, Zhang X, Zwiers FW, Bronaugh D (2013) Climate extremes indices in the CMIP5 multimodel ensemble: part 1. Model evaluation in the present climate. J. Geophys. Res.: Atmospheres 118: 1716–1733

Sillmann J, Kharin V, Zwiers FW, Zhang X, Bronaugh D (2013) Climate extremes indices in the CMIP5 multimodel ensemble: part 2. Future climate projections. J. Geophys. Res.: Atmospheres 118:1716–1733

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker D, Duda MG, Huang X-Y, Wang W, Powers JG (2008) A Description of the Advanced Research WRF Version 3. Technical Note NCAR/

TN–475+STR, University Corporation for Atmospheric Research. https://doi.org/10.5065/D68S4MVH

Tapiador FJ, Navarro A, Moreno R, Sánchez JL, García-Ortega E (2020) Regional climate models: 30 years of dynamical downscaling. Atmos Res 235:104785. https://doi.org/10.1016/j.atmosres.2019.104785

Thompson V, Kennedy-Asser AT, Vosper E, Eunice Lo YT, Huntingford C, Andrews O, M, C, Hegerl GC, Mitchell D (2022) The 2021 western north america heat wave among the most extreme events ever recorded globaly. Science Advance 8(18). https://doi.org/10.1126/sciadv.abm6860

Tomassini L, Jacob D (2009) Spatial analysis of trends in extreme precipitation events in high-resolution climate model results and observations for germany. J. Geophysical Reseach 114:12113. https://doi.org/10.1029/2008JD010652

Xu Z, Han Y, Yang Z (2019) Dynamical downscaling of regional climate: A review of methods and limitations. Sci China Earth Sci 62(2):365–375. https://doi.org/10.1007/s11430-018-9261-5

Zhang X, Alexander L, Hegerl GC, Jones P, Tank AK, Peterson TC, Trewin B, Zwiers FW (2010) Indices for monitoring changes in extremes based on daily temperature and precipitation data. WIREs Clim Change 2(6):851–870. https://doi.org/10.1002/wcc.147

Zollo AL, Rillo V, Bucchignani E, Montesarchio M, Mercogliano P (2016) Extreme temperature and precipitation events over italy: assesment of high-resolution simulations with COSMO-CLM and future scenarios. Int J Climatol 36:987–1004. https://doi.org/10.1002/joc.4401